

Alternative Activation of Macrophages Is Accompanied by Chromatin Remodeling Associated with Lineage-Dependent DNA Shape Features Flanking PU.1 Motifs

Mei San Tang,^{*,1} Emily R. Miraldi,^{†,‡,§} Natasha M. Girgis,^{*} Richard A. Bonneau,^{¶,||} and P'ng Loke^{*,||,#}

IL-4 activates macrophages to adopt distinct phenotypes associated with clearance of helminth infections and tissue repair, but the phenotype depends on the cellular lineage of these macrophages. The molecular basis of chromatin remodeling in response to IL-4 stimulation in tissue-resident and monocyte-derived macrophages is not understood. In this study, we find that IL-4 activation of different lineages of peritoneal macrophages in mice is accompanied by lineage-specific chromatin remodeling in regions enriched with binding motifs of the pioneer transcription factor PU.1. PU.1 motif is similarly associated with both tissue-resident and monocyte-derived IL-4-induced accessible regions but has different lineage-specific DNA shape features and predicted cofactors. Mutation studies based on natural genetic variation between C57BL/6 and BALB/c mouse strains indicate that accessibility of these IL-4-induced regions can be regulated through differences in DNA shape without direct disruption of PU.1 motifs. We propose a model whereby DNA shape features of stimulation-dependent genomic elements contribute to differences in the accessible chromatin landscape of alternatively activated macrophages on different genetic backgrounds that may contribute to phenotypic variations in immune responses. *The Journal of Immunology*, 2020, 205: 1070–1083.

Macrophage activation is a process by which macrophages transition from a resting state to adopt different phenotypes in response to specific external stimuli that can either be danger signals or homeostatic and metabolic signals (1). The macrophage activation process is accompanied by changes in transcriptional activities and histone modifications genome-wide, orchestrated by combinatorial actions of different transcription factors (TFs) that include lineage-determining TFs (such as PU.1) and stimulus-dependent TFs (such as the STAT and IRF proteins) (1–6). However, such molecular events have almost exclusively been described for bone marrow-derived macrophages (BMDMs) in response to TLR signaling, which is often used as a reductionist model to mimic type 1 immune

response to acute infections that gives rise to classically activated macrophages.

In contrast, alternatively activated macrophages (AAMs) induced by type 2 cytokines, such as IL-4 and IL-13, adopt a distinct phenotype that can promote helminth expulsion and limit tissue damage during helminth infection (7–10). Notably, different lineages and phenotypes of tissue macrophages will respond differently to IL-4 stimulation in vivo. We have previously demonstrated that macrophages of tissue-resident and monocytic origins are phenotypically different following IL-4 stimulation (11, 12). Hence, a key question is how and why do different types of macrophages respond differently to the same stimuli? In this study, we expand on these macrophage-specific differences by

^{*}Department of Microbiology, New York University Grossman School of Medicine, New York, NY 10016; [†]Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229; [‡]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229; [§]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267; [¶]Department of Biology, Center for Genomics and Systems Biology, New York University, New York, NY 10003; ^{||}Simons Center for Data Analysis, Simons Foundation, New York, NY 10011; and [#]Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20814

¹Current address: Department of Pathology and Immunology, Washington University in St. Louis, St. Louis, MO.

ORCIDs: 0000-0002-0232-3116 (E.R.M.); 0000-0002-6211-3292 (P.L.).

Received for publication March 16, 2020. Accepted for publication May 17, 2020.

This work was supported in part by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health; National Institute of Allergy and Infectious Diseases, National Institutes of Health Grants AI093811 and AI094166 (to P.L.); National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health Grant DK103788 (to P.L.); the American Association of Immunologists (to M.S.T.); and the Vilcek Foundation (to M.S.T.).

Conceptualization, M.S.T. and P.L.; methodology, M.S.T., E.R.M., N.M.G., R.A.B., and P.L.; formal analysis, M.S.T.; investigation, M.S.T. and N.M.G.; writing – original draft, M.S.T. and P.L.; writing – review and editing, E.R.M. and P.L.; visualization, M.S.T.; data curation, M.S.T.; supervision, R.A.B. and P.L.; funding acquisition, P.L. All authors read and approved of the final draft.

The sequencing data presented in this article have been submitted to the National Center for Biotechnology Information's Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE116108, GSE116107, and GSE116105.

Address correspondence and reprint requests to Dr. P'ng Loke, 9000 Rockville Pike, Room B1-05, Building 4, National Institutes of Health, Bethesda, MD 20892. E-mail address: png.loke@nih.gov

The online version of this article contains supplemental material.

Abbreviations used in this article: AAM, alternatively activated macrophage; AAM^{mono}, alternatively activated F4/80^{int}CD206⁺ macrophage derived from Ly-6C^{hi} inflammatory blood monocyte; AAM^{res}, alternatively activated F4/80^{hi}CD206⁺ macrophage of embryonic origin; ATAC-seq, assay for transposase-accessible chromatin with sequencing; BMDM, bone marrow-derived macrophage; CEL-seq2, cell expression by linear amplification and sequencing 2; ChIP-seq, chromatin immunoprecipitation sequencing; FDR, false discovery rate; FIMO, Find Individual Motif Occurrences; MAPQ, mapping quality; MHCII, MHC class II; M^{mono}, F4/80^{int}CD206⁺ macrophage derived from Ly-6C^{hi} inflammatory blood monocytes; M^{res}, nonstimulated F4/80^{hi}CD206⁺ macrophage of embryonic origin; NYU, New York University; PCA, principal component analysis; qPCR, quantitative PCR; rlog, regularized logarithmic; SNP, single-nucleotide polymorphism; TF, transcription factor; UCSC, University of California, Santa Cruz.

This article is distributed under The American Association of Immunologists, Inc., [Reuse Terms and Conditions for Author Choice articles](#).

Copyright © 2020 by The American Association of Immunologists, Inc. 0022-1767/20/\$37.50

characterizing changes in accessible chromatin landscape following IL-4 stimulation of these different types of macrophages. The effects of IL-4 on BMDMs have been documented, particularly highlighting the reduced response to IFN- γ after IL-4 stimulation that is mediated by the action of TFs such as the PPAR γ /RXR heterodimer and STAT6 (13–15). However, the effects of IL-4 on the chromatin of macrophages from different cellular lineages *in vivo* have yet to be carefully investigated.

Chromatin accessibility changes have mostly been associated with the cellular differentiation process, but we find that IL-4 stimulation alone can give rise to new accessible regions in terminally differentiated peritoneal macrophages. These IL-4-induced regions are dependent on the cellular origin of macrophages and are enriched with binding motifs of PU.1. Leveraging the genetic variations between C57BL/6 and BALB/c mice to perform mutational studies in the accessibility regions remodeled in response to IL-4, we propose and tested a model whereby DNA shape features in the flanking regions to PU.1 motifs contribute to differences in the accessible chromatin landscape of AAMs on different genetic backgrounds. Hence, DNA shape features may be an unappreciated contributor to the remodeling of chromatin in response to external stimuli in macrophages.

Materials and Methods

Experimental methods

Mice. Wild-type C57BL/6 mice were purchased from The Jackson Laboratory and bred onsite for the first set of experiments that compared the effect of IL-4 stimulation on the accessible chromatin profiles from nonstimulated F4/80^{hi}CD206[−] macrophages of embryonic origin (M^{res}), alternatively activated F4/80^{hi}CD206[−] macrophages of embryonic origin (AAM^{res}), F4/80^{int}CD206⁺ macrophages derived from Ly-6C^{hi} inflammatory blood monocytes (M^{mono}), and alternatively activated F4/80^{int}CD206⁺ macrophages derived from Ly-6C^{hi} inflammatory blood monocytes (AAM^{mono}). For experiments directly comparing AAMs of C57BL/6 and BALB/c backgrounds, mice of both strains were purchased from The Jackson Laboratory and immediately used for experiments. Mice were age (7–8 wk of age)– and gender-matched. IL-4/anti-IL-4 mAb complexes (IL-4c) were prepared as described previously (16). To induce AAM^{res}, mice were injected i.p. with IL-4c on days 0 and 2. Mice were also treated with 4% thioglycollate alone (to induce M^{mono}) or in combination with IL-4c to induce AAM^{mono} (12). All animal procedures were approved by the New York University (NYU) Institutional Animal Care and Use Committee under protocol numbers 131004 and 130504.

Peritoneal cell isolation and cell sorting. Peritoneal cells were isolated by washing the peritoneal cavity twice with cold PBS 1 \times . Peritoneal exudate were then treated with ACK Lysis Buffer to lyse RBCs and washed once with PBS. Cells were then resuspended to single-cell suspensions for staining with fluorescently conjugated Abs at 1:100 dilutions, unless otherwise noted. Abs were diluted using 2% FBS. Cells were stained with one of either LIVE/DEAD Blue (Invitrogen) or LIVE/DEAD Near-IR (Invitrogen), blocked with 4 μ g/ml anti-CD16/32 (2.4G2; Bio X Cell) and stained with anti-CD11b Pacific Blue (M1/70; BioLegend), F4/80 PE-Cy7 (BM8; BioLegend), CD206 allophycocyanin (C068C2; BioLegend), Siglec-F PE (E50-2440; BD Biosciences), CD3 PE (145-2C11; BioLegend), CD19 PE (6D5; BioLegend), CD49b PE (DX5; BioLegend), Ly-6G (1A8; BioLegend), PD-L2 (PerCP-Cy55; Miltenyi Biotec; diluted at 1:20), and MHC class II (MHCII) (APC-Cy7; BioLegend). Cells were gated on singlet, live, Dump-negative (CD3[−], CD19[−], DX5[−], Siglec-F[−], Ly-6G[−]), CD11b⁺, then subsequently gated on their M^{res} and AAM^{res} (F4/80^{hi}, CD206[−]) or M^{mono} and AAM^{mono} (F4/80^{int}, CD206⁺) phenotype. Cell surface expression of PD-L2 and MHCII were acquired for analysis. Cells were sorted using 100- μ m nozzle into FBS, on either BD FACSAria II or SONY HAPS1, depending on instrument availability.

Assay for transposase-accessible chromatin with sequencing Assay for transposase-accessible chromatin with sequencing (ATAC-seq) was performed as described by Buenrobro et al. (17). Fifty thousand FACS-purified cells per sample were spun down at 400 g for 5 min at 4°C and washed once with 50 μ l of cold PBS. Cells were lysed with 50 μ l of lysis buffer (10 mM Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM MgCl₂, and 0.1% IGEPAL CA-630) and immediately spun down at 1500 rpm for 10 min at 4°C. The isolated cell nuclei were then incubated for 30 min at 37°C with 50 μ l of

transposase reaction, which contained 25 μ l of Tagment DNA buffer (Illumina), 2.5 μ l of Tagment DNA enzyme (Illumina), and 22.5 μ l of nuclease-free water. The transposed DNA was immediately purified using the Qiagen MinElute PCR Purification Kit (QIAGEN) following the manufacturer's guide and eluted at 10 μ l volume. PCR amplification of the transposed DNA was done using a low-cycle number protocol and with primers published by Buenrobro et al. (17). Each PCR mixture contained of 25 μ l of New England Biolabs 2 \times PCR Mix, 2.5 μ l of 25 μ M forward primer (Primer Ad1_noMX), 2.5 of 25 μ M reverse-barcode primer, 0.3 μ l of 100 \times SYBR Green (Invitrogen), and 10 μ l of transposed DNA. PCR was carried out with the cycling protocol: 72°C for 5 min and 98°C for 30s, followed by five cycles of 98°C for 10 s, 63°C for 30 s, and 72°C for 1 min. The reaction was held at 4°C after the fifth cycle. A side quantitative PCR (qPCR) was set up using the PCR product from these five cycles of amplification. Each qPCR mixture contained 5 μ l of New England Biolabs 2 \times PCR Mix, 0.25 μ l of 25 μ M forward primer, 0.25 μ l of 25 μ M reverse-barcode primer, 0.06 μ l of 100 \times SYBR Green, 4.44 μ l of nuclease-free water, and 5 μ l of the PCR-amplified product. qPCR was carried out using the cycling protocol: 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 63°C for 30 s, and 72°C for 1 min and plate read. The qPCR amplification plot was then used to calculate the additional number of cycles needed for the PCR to achieve maximum amount of product without going into saturation. Each sample was amplified for a total of 14–16 cycles. The amplified libraries were then purified using Qiagen MinElute PCR Purification Kit following the manufacturer's guide and eluted at 20 μ l volume. Libraries were sequenced on the HiSeq 2000 with 2 \times 50 cycles and for an average of 50 million paired-end reads per sample. We performed the IL-4c stimulation experiment twice and generated two independent sets of libraries to obtain an optimal number of biological replicates for each macrophage population. The two independent sets of libraries are referred to in this study as Run1 and Run2, respectively. ATAC-seq libraries for C57BL/6 and BALB/c AAMs were generated using the same protocol and sequenced in a single run.

Transcriptional profiling of BALB/c and C57BL/6 AAMs. One hundred thousand cells were sorted per sample as described above. FACS-purified cells were spun down and washed once with PBS before lysis with 350 μ l of Buffer RLT from the RNeasy Mini Kit (QIAGEN). RLT lysates were homogenized by 1 min of vortexing and were immediately stored at −80°C until RNA isolation. RNA was isolated using the RNeasy Mini Kit (QIAGEN) based on the manufacturer's protocol, with an additional DNase digestion step using the RNase-free DNase set (QIAGEN). Transcriptional profiling was done using the cell expression by linear amplification and sequencing 2 (CEL-seq2) protocol (18), and library preparation was performed at NYU School of Medicine Genome Technology Center core facility. CEL-seq2 libraries were sequenced on the HiSeq 4000 with 2 \times 50 cycles. Although CEL-seq2 was originally developed as a single-cell assay, we used this protocol in this study as a bulk transcriptional profiling assay and use the more commonly used term RNA-seq to describe data generated from this assay.

Bioinformatics and computational methods

ATAC-seq processing. Raw ATAC-seq reads were aligned to the reference mouse genome mm9 using bowtie2 (v2.2.9) (19), with the parameters maxin 2000 and local while keeping all other parameters at default settings. To keep only highly unique alignments, reads with mapping quality (MAPQ) score <30 were removed. We further removed all duplicate reads, as well as reads mapping to mitochondrial DNA and chromosome Y. Read filtering steps were done using the suite of tools from samtools (v1.2 and v1.3.1) (20), ngsutils (v0.5.9) (21), and picard-tools (<http://broadinstitute.github.io/picard/>; v1.1.1 and v2.8.2). After all filtering steps, reads were merged across all replicates from the same macrophage population. This resulted in a median depth of 15,235,324 reads per macrophage population in Run 1 and 9,865,310 reads per macrophage population in Run 2. For visualization of accessibility reads on the Integrative Genomics Viewer, we merged reads from the same macrophage population across samples from both runs, generated tiled data format files using Integrative Genomics Viewer tools and finally normalized the merged reads to reads per million (22).

Identification of accessible chromatin regions. We used the merged reads for each macrophage population to identify accessible chromatin regions, using the PeakDECK (v1.1) peak calling algorithm, which measures signal density from randomly sampled bins genome-wide before generating a dataset-specific probability distribution to identify regions with significant signal enrichment (23). We ran PeakDECK using sampling bins that consist of a 75-bp central bin (−bin) and a 10,000 bp background bin (−back). Sampling along the genome was done in steps (−STEP) of 25 bp and the background probability distribution was generated using 100,000

randomly selected sites (-npBack). Significance was defined using a p value < 0.0001 , and regions with significant p values were defined as a “peak” (i.e., an accessible chromatin region). Peak calling was done independently on libraries generated from Run 1 and Run 2.

Generation of a union set of accessible chromatin regions. We next counted the number of reads present at each accessible region to analyze the ATAC-seq data using quantitative approaches downstream. To do this, we first generated a set of consensus peaks across the dataset by taking the union of peaks called from each macrophage population. Peaks were merged if overlapping by 1 bp or more. The number of reads at each peak within the union peak sets were then counted for each sample. Finally, each peak was centered ± 100 bp on its summit, defined as the position with maximum pile up of reads. Recentering on peak summits was performed as this should coincide with the binding event of a TF within an accessible chromatin region. We implemented the read counting and peak summit recentering steps directly using the `dba.count` function from the Bioconductor package `DiffBind` (version 1.14.2) (24). The final count matrix, which consisted of 61,713 peaks, was used for downstream analyses.

Quantitative ATAC-seq analysis. ATAC-seq read counts were transformed using the regularized logarithmic (rlog) transformation as implemented in the Bioconductor package `DESeq2` (25). To manage batch effect from the two separate libraries, we first modeled the rlog accessibility read counts to the batch variable using a linear model and subtracted out the coefficient contributed by the batch variable—this was implemented directly using the `removeBatchEffect` function in `limma` (26). We next chose a set of 30,856 regions with high variance, using the `varFilter` function in the `genefilter` package with default parameters, which keeps only features with variance interquartile range > 0.5 (27). We performed principal component analysis (PCA) using the batch-subtracted rlog read counts of these regions with high variance using the `prcomp` function in R.

To identify IL-4-dependent accessible regions, we directly compared the ATAC-seq profiles of IL-4-stimulated macrophages to their reference nonstimulated macrophages using a differential analyses workflow directly implemented through `DESeq2`. We fit the negative binomial model in `DESeq2` using the raw accessibility reads from all 61,713 regions, with the model \sim Batch + Population, in which Batch is a variable describing if a sample belonged in Run 1 or Run 2, whereas Population is a variable describing if the sample is M^{res} , AAM^{res} , M^{mono} , or AAM^{mono} . IL-4 dependency was defined using a significance threshold of false discovery rate (FDR) of 10%. To visualize IL-4-dependent regions, we scaled the batch-subtracted rlog read counts of these IL-4-dependent regions by z -score transformation and next performed k -means clustering on these scaled, rlog-transformed reads ($K = 4$). The clustered matrix was visualized as a heatmap.

Comparison of sequence properties between constitutively accessible and IL-4-induced regions. Identification of constitutively accessible and IL-4-induced regions. To define a set of constitutively accessible regions, we used only peaks from M^{res} and M^{mono} , respectively, that were identified in both Run 1 and Run 2. This resulted in 8061 constitutively accessible regions in M^{res} and 14,045 constitutively accessible regions in M^{mono} . IL-4-induced peaks were defined using the differential analysis outlined above. All region overlap analyses throughout this study were performed using the `intersect` function from the `BEDTools` suite (28), and overlaps were defined as any regions overlapping by at least 1 bp, unless otherwise noted.

Genomic elements enrichment analysis. We downloaded genome-wide annotations of five different genomic elements (promoter, start exon, coding exon, end exon, and intron) from the University of California, Santa Cruz (UCSC) Known Gene database for mm9 (29). We defined promoter elements as the 200-bp-region upstream of a transcriptional start site. We next assigned each of the 61,713 accessible regions in our dataset to a unique genomic element label. When an accessible region overlapped two different types of genomic elements, we assigned it to the element with higher number of overlapping base pair. Finally, any chromatin regions not assigned to one of these five genomic elements were labeled as intergenic. To determine the enrichment levels of a particular type of genomic element G within a given set of accessible regions A (either constitutively accessible or IL-4-induced regions in AAM^{res} or AAM^{mono}), we used the binomial cumulative probability distribution, $b(x; n, p)$, where x = number of success, n = number of trials, and p = background probability of success. We used the `pbinom` function in R. We defined x to be the number of accessible regions in A that were labeled as the genomic element G that was being tested, n to be the total number of genomic elements G detected in our combined dataset and p to be the proportion of the accessible region A to the total 61,713 accessible regions. This then gave the enrichment levels of G in A , relative to all the accessible regions identified across the different macrophage populations.

G/C content analysis. To calculate percentage G/C, we first used the `hgGcPercent` function from the `kentTools` suite (v20170111, UCSC Genome Bioinformatics Group; <https://github.com/ucscGenomeBrowser/kent>) to quantitate the number of G or C bases in each accessible region. This value was next normalized using the length of the accessible region. CpG island track was downloaded from the UCSC Genome Annotation Database for mm9 (<http://hgdownload.soe.ucsc.edu/goldenPath/mm9/database/>). Enrichment levels of CpG island in a given set of accessible regions A was based on the binomial cumulative probability as described above, in which x = number of accessible regions in A that overlapped a CpG island, n = number of CpG island in the total dataset of 61,713 regions, and p = proportion of A to the total 61,713 regions.

Calculation of distance to IL-4-induced genes. IL-4-induced genes were first identified for AAM^{res} and AAM^{mono} from the microarrays generated by Gundra and Girgis et al. (12), using the linear model and empirical Bayes statistics as implemented in `limma`, with genes significantly induced by IL-4 defined using the thresholds FDR 10% and \log_2 fold change > 1 . The distance between each accessible region and its closest IL-4-induced gene body was calculated using the closest function in `BEDTools`.

TF motif analysis. Sequences of IL-4-induced regions were fetched using the `BEDTools` `getfasta` function for TF motif analysis with the `MEME` Suite tools (30). Whole genome fasta file for mm9 was downloaded from the Illumina igenome database (https://support.illumina.com/sequencing/sequencing_software/igenome.html). Background file was generated using the function `fasta-get-markov` in `MEME`, based on the total 61,713 accessible regions at a Markov model order of 3. TF motif databases (which included mouse and human TF motifs) were curated as described in Ref. 31. We performed de novo motif discovery by running `MEME` (as part of `MEMEChIP`, which randomly sampled 600 sequences) with the following parameters: `-mod zoops -nmotifs 3, -minw 6, and -maxw 30`. Overrepresentation analysis to identify macrophage-specific TF motifs was performed by running `HOMER` (32) using sequences from the opposing macrophages as background sequences (i.e., to identify TF motifs specific to AAM^{res} , sequences of IL-4-induced regions from AAM^{mono} were used as background sequences) and the parameters: `-mask, -size 8,10,12,16, -mset vertebrates, and nlen 3`. For motifs from each macrophage lineage, we combined all three de novo motifs discovered by `MEME` and motifs with enrichment $\log_2 p$ values < -15 by `HOMER` for clustering analysis (we used the known motifs output from `HOMER`). This resulted in 14 motifs for AAM^{res} and AAM^{mono} , respectively. Clustering was done using `STAMP` (33, 34), with the frequency matrices of motifs and the following default parameters: column comparison metric - Pearson correlation coefficient, alignment method - ungapped Smith-Waterman, tree-building algorithm - UPGMA, and multiple alignment strategy - iterative refinement.

Comparisons between predicted PU.1 motif and chromatin immunoprecipitation sequencing-defined PU.1 binding sites. PU.1 chromatin immunoprecipitation sequencing (ChIP-seq) regions identified in thioglycollate-induced macrophages, generated from two different experiments, were downloaded as browser extensible data (BED) files that had been directly deposited in the Gene Expression Omnibus (GSM1131238 and GSM1183968) (5, 35). A set of 55,386 reproducible PU.1 binding sites were defined by intersecting these two sets of PU.1 ChIP-seq regions. We used the PU.1 motif discovered de novo from all the constitutively accessible regions in M^{mono} and ran `Find Individual Motif Occurrences` (FIMO) (36) to identify all PU.1 motif sites from M^{mono} using a p value threshold of 0.0001 and the background file generated as described above. Because the published PU.1 ChIP-seq regions were of 200-bp length, we extended the predicted PU.1 motifs from M^{mono} by ± 100 bp to match the comparison. Overlapping rate was calculated as total number of predicted PU.1 motifs from M^{mono} overlapping a reproducible PU.1 ChIP-seq region/total number of predicted PU.1 motifs from $M^{\text{mono}} \times 100\%$.

PU.1 motif analysis. We identified IL-4-induced PU.1 motif sites by performing motif scanning with FIMO using the PU.1 motif discovered de novo from the IL-4-induced peaks of AAM^{res} and AAM^{mono} , respectively. FIMO was run with a p value threshold of 0.0001 (as part of `MEMEChIP`). Motif scores were also calculated as part of FIMO.

The detected IL-4-induced PU.1 motif sites were then extended ± 25 bp using the `BEDTools` `slop` function. These PU.1 motif sites ± 25 bp are hereafter referred to as “PU.1 regions.” To identify potential cofactors for PU.1, these PU.1 regions were specifically subjected to motif scanning by FIMO, with a background model that was based on the 55,386 PU.1 ChIP-seq peaks described above and at a Markov model order of 3. To determine which of these detected motifs were macrophage specific, the odds ratio of a motif being detected in the PU.1 regions of AAM^{res} versus AAM^{mono} were calculated. The p value was determined using two-sided Fisher test, with the null hypothesis of a motif being equally likely to be detected in the IL-4-induced PU.1 regions of AAM^{res} and AAM^{mono} (i.e., \log_2 odds

ratio of 0). When multiple motifs of the same TF were present in the database, we used the motif that was most frequently detected in PU.1 regions for odds ratio calculations. TFs from mouse and human were kept as separate analyses, but included in the same visualization. To account for the multiple hypotheses testing performed over 633 mouse TFs and 835 human TFs, the Benjamini–Hochberg procedure was used to perform p value adjustment by calculating the FDR and significance threshold was set at FDR 10%. Hence, statistically significant TFs were macrophage specific, with \log_2 odds ratio >0 , indicating AAM^{res} specificity and \log_2 odds ratio <0 indicating AAM^{mono} specificity. For visualization of significant results, TFs were summarized at the family level as defined in Ref. 37 (Fig. 3B). The maximum absolute \log_2 odds ratio of the family was visualized and the specific TF with the maximum absolute \log_2 odds ratio value was stated in parenthesis. Where TF family annotation was not available, the \log_2 odds ratio of the specific TF itself was used.

DNA shape features of PU.1 motifs were analyzed using the DNashape algorithm (24, 38, 39) for four different DNA shape configurations at single-nucleotide resolution. Sequences on the antisense strand were reverse complemented prior to DNA shape prediction.

ATAC-seq processing from C57BL/6 versus BALB/c experiment

Raw ATAC-seq reads were aligned to the mm10 mouse reference genome (for compatibility with available BALB/c sequence variant data, see details below). Reads with MAPQ <30 , as well as duplicate and mitochondrial reads, were all removed before peak calling. Peak calling in PeakDeck was done using the same parameters in the IL-4c stimulation experiment, except for a p value threshold of 0.00001. A consensus count matrix consisting of 40,981 accessible regions from all samples was generated as described above using DiffBind and used to identify strain-specific regions using the differential analysis workflow implemented through DESeq2. We fit the negative binomial model using \sim Strain + CellType + Strain:CellType, where Strain is one of either C57BL/6 or BALB/c and CellType is one of either AAM^{res} or AAM^{mono}, and extracted strain-specific regions from the following comparisons: 1) BALB/c versus C57BL/6 in AAM^{res}, and 2) BALB/c versus C57BL/6 in AAM^{mono}. Significance was defined at a threshold of FDR 10%.

Characterizing BALB/c variants at PU.1 motif regions

BALB/c-specific sequence variants, including single nucleotide polymorphisms (SNPs) and insertion/deletion (indels), were downloaded as variant call format (VCF) files from the Sanger Mouse Project (<ftp://ftp-mouse.sanger.ac.uk/>). Variants that were not labeled “PASS” under the Filter flag were removed using vcfTools (40). These VCF files were then converted to the BED file format using the vcf2bed function from bedops (41) for overlap analysis with the IL-4-induced PU.1 regions. Because these BALB/c variants were identified on the mm10 genome, we lifted over IL-4-induced PU.1 regions defined initially on the mm9 coordinates to the mm10 coordinates. PU.1 motif sites that directly overlapped sequence variants were identified using the BEDTools intersect function. To define PU.1 motif sites with flanking variants, we used the BEDTools closest function to calculate distances between a PU.1 motif and its closest BALB/c variant and identified motifs with a closest variant at an absolute distance of <25 bp, but not overlapping a variant directly. We also limited the DNA shape prediction to SNPs only, to avoid the shift in position due to the insertions/deletions.

Generation of BALB/c whole genome fasta

To construct a BALB/c whole genome fasta for DNA shape prediction, we took the list of BALB/c variants filtered from above and used the vcf2diploid tool (42) to integrate these BALB/c variants into the mm10 reference genome. To annotate PU.1 regions on the BALB/c coordinates, we used the accompanying chain file generated from the vcf2diploid tool in the above step to lift over the PU.1 regions from mm10 coordinates to BALB/c coordinates. The mm10 and BALB/c whole genome fasta and the PU.1 regions on the corresponding C57BL/6 and BALB/c coordinates were then used with the getfasta function in BEDTools to fetch the sequences of specific IL-4-induced PU.1 regions with flanking SNPs for shape prediction with DNashape algorithm.

PU.1 and cofactor ChIP-seq analysis

The PU.1 and cofactor ChIP-seq analyses were performed on data from Link et al. (43). BED files were directly downloaded from the Gene Expression Omnibus (GSE109965). We first identified analyzed data sets from nonstimulated BMDMs of C57BL/6 background. TF ChIP-seq that were

analyzed included PU.1, CEBP, cJUN, RUNX, and USF. Only TF binding sites that were accessible (as defined by ATAC-seq) and contained H3K27ac histone modifications were included in the analysis. PU.1 binding sites that were not cobound by other cofactors were defined as PU.1 ChIP-seq peaks that did not overlap with ChIP-seq peaks of CEBP, cJUN, RUNX, or USF, using the intersect function in BEDtools. This resulted in 673 PU.1 ChIP-seq peaks.

To identify PU.1 motif sites from these 673 PU.1 ChIP-seq peaks, we first derived the PU.1 motif from nonstimulated BMDMs by performing MEME on all the PU.1 ChIP-seq peaks (a random background at Markov order 3). The de novo PU.1 motif was then used to run FIMO on the 673 PU.1 ChIP-seq peaks that did not overlap with other TFs (p value threshold of 0.0001; background was based on the union set of ATAC-seq peaks in BMDMs at Markov order 3). Finally, we defined strain-specific PU.1 binding sites as PU.1 motifs in peaks with at least 4-fold reduction in SPRET/Ei PU.1 ChIP-seq reads compared with C57BL/6 PU.1 ChIP-seq reads.

Identification of predictors for strain-specific and strain-common regions. To identify features associated with the 922-strain-specific and 4776-strain-common SNP-carrying accessible regions, we first assessed each of these accessible regions for the following characteristics: EuclideanDistanceProT = amount of shape change induced by the BALB/c SNP(s) in the ProT shape configuration. EuclideanDistanceRoll = amount of shape change induced by the BALB/c SNP(s) in the roll shape configuration. EuclideanDistanceMGW = amount of shape change induced by the BALB/c SNP(s) in the MGW shape configuration. GCPercentB6 = % G/C calculated using C57BL/6 sequences. CpGB6 = a binary variable (yes/no) indicating if the region overlapped a CpG island. GenomicElements = the type of genomic element that the region is labeled as (one of promoter, start exon, end exon, coding exon, intron, or intergenic). SnpFrequency = Number of SNPs carried by the region. IL4Inducibility = a binary variable (yes/no) indicating if the region was IL-4 inducible. We then fit a logistic regression model as follows using the glm function in R, with the parameter family = “binomial”: $\text{Class} \sim \text{EuclideanDistanceProT} + \text{EuclideanDistanceRoll} + \text{EuclideanDistanceMGW} + \text{GCPercentB6} + \text{CpGB6} + \text{GenomicElements} + \text{SnpFrequency} + \text{IL4Inducibility}$. The dependent variable “Class” is a label describing if the region is “strain-specific” or “strain-common.” Significance of predictor was determined by ANOVA for the coefficients of the predictors, using a p value cutoff of <0.05 .

Processing of CEL-seq reads. CEL-seq reads were first demultiplexed using the bc_demultiplex script from <https://github.com/yanailab/CEL-Seq-pipeline> (18). Demultiplexed reads were aligned to the mm10 mouse reference genome using bowtie2 (version 2.2.9). Aligned reads were counted for each gene using a modified htseq-count script (from <https://github.com/yanailab/CEL-Seq-pipeline>) adapted for CEL-seq reads with unique molecular identified. We included only reads with mapping quality score >30 and removed singleton genes. This resulted in a final median read depth of 737,848 reads per sample, covering a median of 11,096 genes per sample. PCA was performed using 7431 genes with high variance, defined using the varFilter function in the genefilter package with default parameters, which keeps only features with variance interquartile range >0.5 . Differential analysis was done using DESeq2 by fitting the negative binomial model using \sim Strain + CellType + Strain:CellType. Significantly differential genes were extracted using a threshold of FDR 10% for the four different comparisons of 1) AAM^{res} versus AAM^{mono} in C57BL/6, 2) AAM^{res} versus AAM^{mono} in BALB/c, 3) BALB/c versus C57BL/6 in AAM^{res}, and 4) BALB/c versus C57BL/6 in AAM^{mono}. Overlapping genes were defined as genes identified as differential in two different comparisons. Pathway enrichment analysis was done using strain-specific genes in AAM^{res} and AAM^{mono}, respectively, through Ingenuity Pathway Analysis with the parameter Organism = Mouse and keeping all other parameters at default settings.

Results

IL-4 stimulation leads to remodeling of open-chromatin landscape in peritoneal macrophages

To examine chromatin remodeling on different types of tissue macrophages, we injected IL-4c into the peritoneal cavity of C57BL/6 mice to induce accumulation of AAM^{res} and compared these with AAM^{mono} in mice injected with IL-4c and thioglycollate (12). We then used ATAC-seq (17) to profile the open-chromatin landscape of these macrophages, in comparison with nonstimulated F4/80^{hi}CD206[−] macrophages of naive mice and F4/80^{int}CD206⁺ macrophages from thioglycollate-treated mice (12).

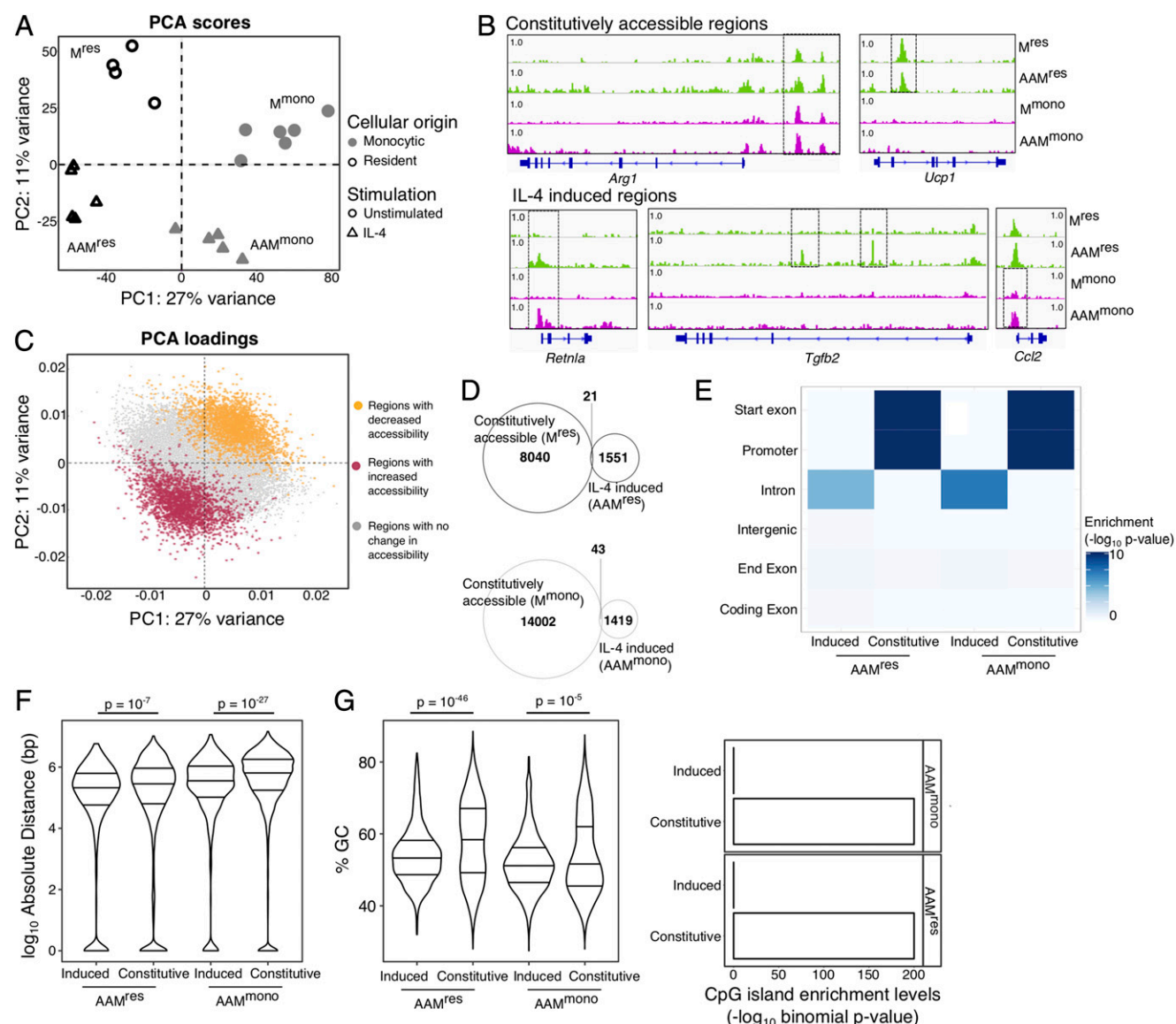


FIGURE 1. IL-4 stimulation leads to remodeling of open chromatin landscape in peritoneal macrophages. **(A)** PCA scores of individual ATAC-seq samples. PCA was performed using rlog-transformed ATAC-seq read counts of 30,856 regions with high variance (only regions with variance interquartile range >0.5 were retained) ($n = 4$ – 6 mice per macrophage population). Data points represent independent biological replicates. **(B)** Genome browser views of representative (boxed) constitutively accessible and IL-4-induced regions. **(C)** The contributions of individual accessible regions to principal components 1 and 2 are represented in the PCA loadings plot. Each data point is color-coded based on the direction of its IL-4 dependency. Hence, IL-4-induced regions (red) are highly associated with IL-4-stimulated macrophages, whereas IL-4-repressed regions (yellow) are highly associated with nonstimulated macrophages. **(D)** Venn diagrams indicating the number of constitutively and IL-4 induced accessible regions in (top) resident and (bottom) monocytic macrophages. We compared **(E)** enrichment levels for different types of genomic elements, **(F)** distance from a closest IL-4-induced gene, and **(G)** G/C content between constitutively accessible and IL-4-induced regions in AAM^{mono} and AAM^{res} , respectively. G/C content information is represented in two different ways: percentage of G/C bases in an accessible region (G, left panel) and CpG island enrichment for a given group of accessible regions (G, right panel). Number of IL-4-induced regions = 1572 in AAM^{res} and 1462 in AAM^{mono} ; number of constitutively accessible regions = 8061 in AAM^{res} and 14,045 in AAM^{mono} . Enrichment p values are from binomial test, whereas two-class comparison p values are from two-sided Mann–Whitney test.

The overall differences in accessible chromatin landscape (a total of 61,713 open-chromatin regions) could be attributed mainly to the cellular origin of macrophage (27% of total variance), but alternative activation by IL-4 also altered the accessible chromatin profiles (Fig. 1A, 1B). *Arg1* and *Ucp1*, which are known to be IL-4 inducible (12), had constitutively accessible chromatin regions, whereas *Retnla*, another IL-4 inducible gene, had chromatin regions that gained accessibility in response to IL-4 (Fig. 1B). This IL-4-induced chromatin remodeling process can also be specific to either tissue-resident or monocyte-derived macrophages at certain regions (e.g., regions adjacent to the loci of

Tgfb2, *Ccl2*) (Fig. 1B). Of the 61,713 total accessible regions, we identified 1572 regions induced by IL-4 for AAM^{res} and 1462 regions for AAM^{mono} (Supplemental Fig. 1A). IL-4-dependent regions also had the largest contribution to the differences in open-chromatin profiles between nonstimulated and IL-4-stimulated macrophages (Fig. 1C).

The IL-4-induced accessibility regions almost all (99% in AAM^{res} and 97% in AAM^{mono}) gained accessibility from undetectable levels at baseline (Fig. 1D). To determine how the IL-4-induced accessibility regions differed from regions that were constitutively accessible, we made comparisons of several sequence characteristics

between constitutively accessible and IL-4-induced regions (Supplemental Fig. 1B). IL-4-induced regions were more likely to reside in noncoding intronic regions—in tissue-resident macrophages, 781 of 1572 IL-4-induced peaks were intronic, as compared with 2199 of 8061 constitutively accessible regions (binomial enrichment; $p = 2.5 \times 10^{-5}$), whereas in monocyte-derived macrophages, 761 of 1462 IL-4-induced peaks were intronic, as compared with 5282 of 14,045 constitutively accessible regions (binomial enrichment; $p = 8.5 \times 10^{-4}$) (Fig. 1E). In the monocyte-derived macrophages, IL-4-induced regions were overall closer to IL-4-induced genes (two-sided Mann–Whitney U test; $p = 3.4 \times 10^{-25}$ in AAM^{mono}) (Fig. 1F). In tissue-resident macrophages, IL-4-induced regions contained lower G/C content (two-sided Mann–Whitney U test; $p = 8.6 \times 10^{-46}$ in AAM^{res}) (Fig. 1G, left) and were also less likely to overlap with a CpG island (Fig. 1G, right). Hence, IL-4 stimulation can lead to reorganization of the chromatin landscape in terminally differentiated peritoneal macrophages, giving rise to newly accessible regions that have distinct sequence properties when compared with constitutively accessible regions. The stimulation induced accessibility changes are more likely to be in noncoding intronic regions of the genome.

IL-4-induced regions are associated with PU.1, KLF, and AP-1 motifs

Even though both AAM^{mono} and AAM^{res} received the same stimulation within the peritoneal tissue environment, the regions that were remodeled by IL-4 were largely dependent on cellular origin (Fig. 2A). Of all the 2855 IL-4-induced regions, only 179 regions (6% of total IL-4-induced regions) were shared between both AAM^{mono} and AAM^{res} (Supplemental Fig. 1C). Although the IL-4-induced regions from AAM^{mono} and AAM^{res} were largely distinct, the DNA motifs discovered from these distinct regions were grouped into similar families of TFs, which included PU.1, KLF, and the AP-1 family of motifs (Fig. 2B). However, AAM^{mono} had significantly higher number of accessible regions with the AP-1 motif (two-sided Fisher exact test; $p = 9.5 \times 10^{-10}$), whereas AAM^{res} had significantly higher number of accessible regions with the KLF motif (two-sided Fisher exact test; $p = 9.2 \times 10^{-8}$), suggesting the use of different TFs by the two macrophage types during chromatin remodeling upon IL-4-induced alternative activation (Fig. 2C). Such differences were not observed with the PU.1 motifs. We next examined the expression levels of these TF families (KLF versus AP-1) to determine if specific members within each family are differentially expressed between AAM^{res} and AAM^{mono}. Twenty TFs of KLF and AP-1 families were highly expressed in peritoneal macrophages (Supplemental Fig. 2A). These TFs almost all demonstrated lineage-specific expression, both at baseline (clusters 1 and 2) and with IL-4 stimulation (clusters 3 and 4) (Fig. 2D). These results suggest that although KLF and AP-1 family of TFs may have lineage-specific functions, PU.1 is likely important for both macrophage lineages.

We next used an overrepresentation approach to identify additional motifs enriched in the IL-4-induced regions of AAM^{res} or AAM^{mono}. Because there were some overlaps between the motifs discovered by the de novo discovery method and overrepresentation method, we combined the two sets of motifs for a clustering analysis and merged motifs that were redundant. Using this approach, we identified macrophage-specific motifs beyond those from de novo motif discovery (Fig. 2E). The GATA motifs and basic helix-loop-helix (bHLH) motifs were specific to IL-4-induced regions of AAM^{res}. TFs with these binding motifs have been implicated to be important in proliferation of tissue-resident macrophages (44, 45). In contrast, the NFY and STAT motifs were

specific to IL-4-induced regions of AAM^{mono}. These macrophage-specific motifs were only detected in 15–21% of IL-4-induced regions (236 of 1572 IL-4-induced peaks in AAM^{res} and 317 of 1462 IL-4-induced peaks in AAM^{mono}), whereas the PU.1, KLF, and AP-1 motifs discovered by the de novo method were present in ~75% of IL-4-induced peaks in AAM^{res} and AAM^{mono}. Therefore, although there are specific TF motifs enriched in IL-4-induced regions of AAM^{res} (GATA and bHLH) and AAM^{mono} (NFY and STAT), the majority of IL-4-induced regions are enriched for a common set of TF motifs. Hence, differential accessibility of putative TF binding sites alone could not explain the lineage-specific chromatin remodeling patterns for macrophages of different cellular origins. We hypothesized that other sequence features around the PU.1 motifs could be important.

PU.1 motifs in AAM^{res} and AAM^{mono} are associated with macrophage-specific sequence features

PU.1 motif was the most frequently found motif from the IL-4-induced regions across both AAM^{res} (639 of 1572 IL-4-induced regions) and AAM^{mono} (552 of 1462 IL-4-induced regions) (Fig. 2C). We focused on these predicted PU.1 binding sites and further characterized their local sequence features in both types of macrophages. In the absence of ChIP-seq data from our own experiments, we first quantified the accuracy of PU.1 motif prediction with actual PU.1 binding, by comparing our predicted PU.1 motifs from M^{mono} with PU.1 ChIP-seq data in the same cell type (5, 35). A total of 78% of the PU.1 motif sites predicted from M^{mono} in our study (4282 of total 5492 predicted PU.1 motif sites) overlapped with a PU.1 binding site defined by ChIP-seq (Supplemental Fig. 2B). This provided some confidence that most of the PU.1 motifs we were investigating are likely to be directly bound by PU.1.

We next characterized the PU.1 motifs discovered from the IL-4-induced regions of AAM^{res} and AAM^{mono}. Motif score is commonly used as a proxy of TF-DNA binding affinity. When compared with PU.1 motifs from AAM^{res}, PU.1 motifs from AAM^{mono} had significantly lower motif scores and also demonstrated greater variability in their values (Fig. 3A). To identify potential cofactors that could bind in collaboration with PU.1 and contribute to macrophage-specific PU.1 accessibility, we performed motif scanning using sequences from PU.1 motifs \pm 25 bp flanking sequences (Supplemental Fig. 2C). We identified TF motifs that were specific for AAM^{res} versus AAM^{mono} in these PU.1 regions (Fig. 3B). These predicted macrophage-specific cofactors are largely from different families. We observed a greater diversity in TF families enriched around the IL-4-induced PU.1 motifs of AAM^{mono}. These results indicate that PU.1 could function cooperatively with different cofactors to bind different genomic regions depending on the macrophage lineage. However, because the majority of TF families were enriched at similar levels in both AAM^{res} and AAM^{mono}, differential accessibility of putative TF binding sites alone could not explain the macrophage lineage-specific chromatin remodeling patterns. We hypothesized that other sequence features around PU.1 motifs could be important.

Because functional PU.1 binding can be most accurately determined by TF motifs neighboring the PU.1 motif and by the local three-dimensional shape of the DNA at PU.1 binding sites and DNA shape can be a predictor of TF binding pattern (46, 47), we next determined if such shape properties could be associated with differences in IL-4-induced PU.1 regions between AAM^{res} and AAM^{mono}. We computationally predicted four DNA shape configurations (minor groove width, propeller twist, and helical twist and roll) at the IL-4-induced PU.1 regions of AAM^{res} and AAM^{mono} and found that IL-4-induced PU.1 regions of AAM^{res} and AAM^{mono} differed significantly in shape profiles (Fig. 3C). The PU.1 motifs of

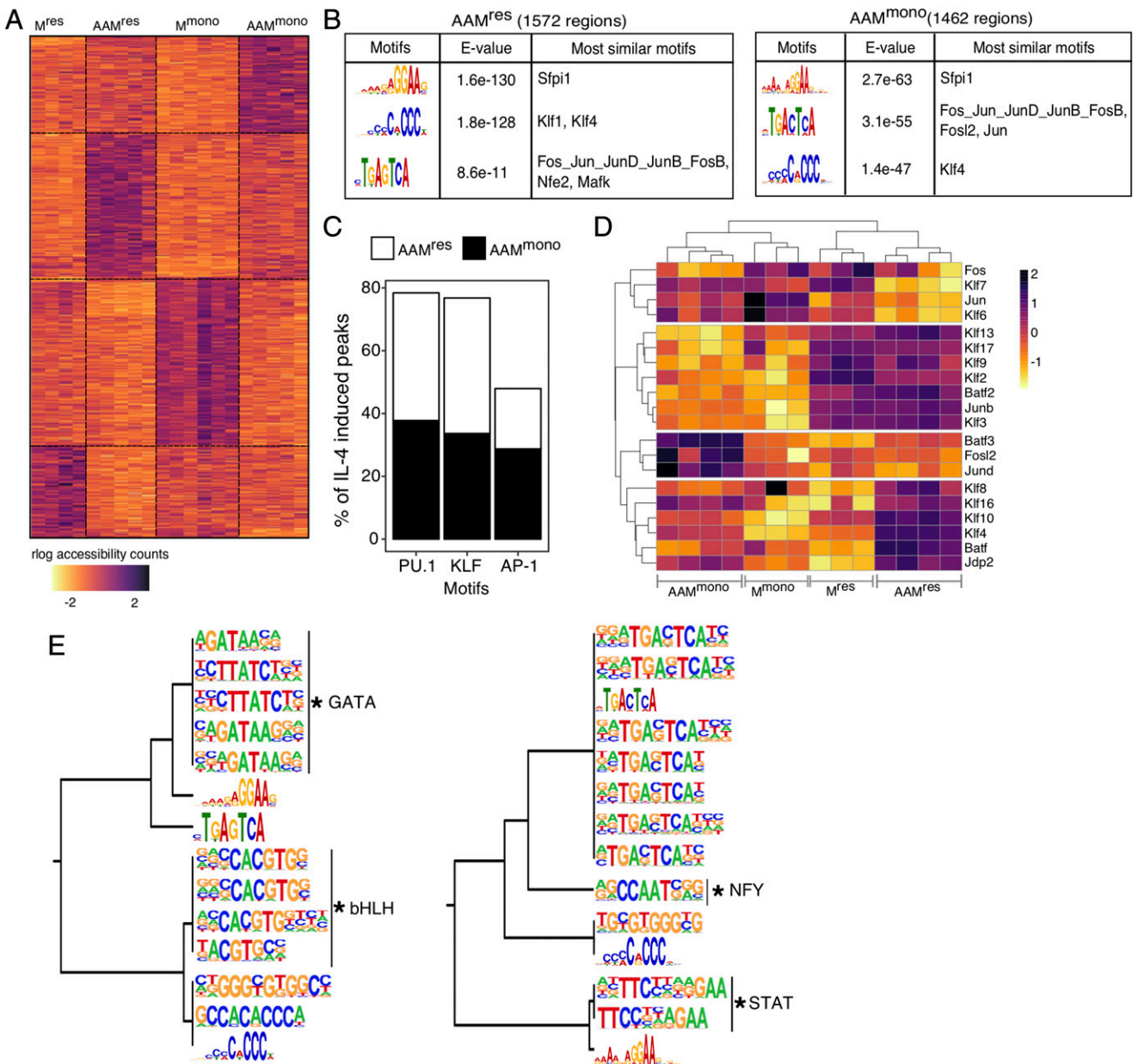


FIGURE 2. IL-4-induced regions are associated with PU.1, KLF, and AP-1 motifs. **(A)** Heatmap visualizing the macrophage-specific IL-4-dependent regions. Each row represents one of the 2855 IL-4-dependent regions and each column a unique sample. Values are rlog-transformed, batch-subtracted read counts, scaled using a z-score transformation for each region. **(B)** Motifs discovered de novo from IL-4-induced regions in AAM^{res} and AAM^{mono} . **(C)** Frequency of IL-4-induced peaks delineated by the presence of de novo PU.1, KLF, and AP-1 motifs. **(D)** Twenty highly expressed TF genes related to the de novo motifs discovered from the IL-4-induced regions. Values are log₂ intensity values of microarrays (12). **(E)** Clustering analysis of de novo motifs and macrophage-specific motifs identified using an overrepresentation approach from the IL-4-induced regions of AAM^{res} (left) and AAM^{mono} (right). Asterisks indicate macrophage-specific motifs uniquely identified via the overrepresentation approach. Only macrophage-specific motifs with log₂ *p* value < -15 are included in this visualization.

AAM^{res} have a more conserved DNA configuration, whereas PU.1 motifs of AAM^{mono} demonstrated greater variability in DNA configuration. These differences were most pronounced in the propeller and helical twist configurations of the eighth base pair in the PU.1 motif (Fig. 3D). These macrophage-specific DNA shape profiles were specific to the PU.1 motif regions, as IL-4-induced KLF motif regions did not display AAM^{res} and AAM^{mono} specific DNA shape profiles (Supplemental Fig. 2D). These results point toward the contribution of local DNA shape on the macrophage lineage-specific pattern of PU.1 motif accessibility and binding. Although the PU.1 motif is the most enriched motif in the IL-4-induced regions for both AAM^{res} and AAM^{mono} , sequence characteristics in the PU.1 motifs differ between the two types of peritoneal macrophages, suggesting

local DNA shape features of PU.1 binding sites might cooperate with distinct coregulators in these different lineages to contribute toward lineage-specific AAMs in response to IL-4 stimulation.

Accessibility of IL-4-induced PU.1 regions can be altered through DNA shape change induced by sequence mutation, without disruption of the PU.1 motif

To determine whether altering the local DNA shape would change accessibility of the PU.1 motifs, we designed a mutation study using natural variants between the C57BL/6 and BALB/c mouse strains, by performing ATAC-seq on AAM^{mono} and AAM^{res} generated from mice of these two different genetics backgrounds. We first defined strain-specific (significant differences

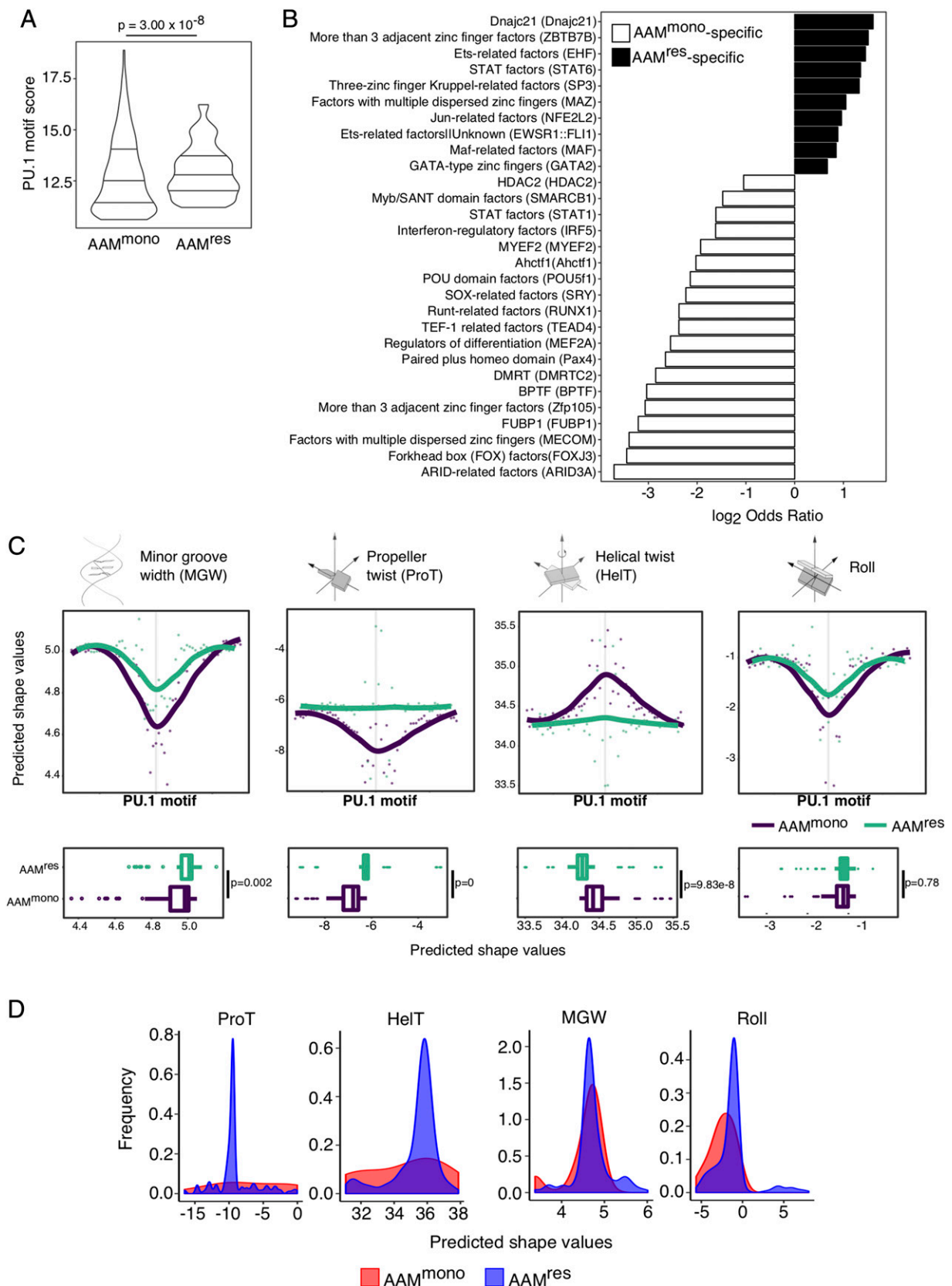


FIGURE 3. PU.1 motifs in AAM^{res} and AAM^{mono} are associated with macrophage-specific sequence features. **(A)** Comparison of PU.1 motif scores derived by FIMO in IL-4-induced regions of AAM^{res} versus AAM^{mono}, with horizontal lines in the violin plots representing values at 25th, 50th, and 75th percentiles. The p value is from a two-sided Mann–Whitney U test. Number of IL-4-induced regions = 1572 in AAM^{res} and 1462 in AAM^{mono}. **(B)** Macrophage-specific TF motifs found within IL-4-induced PU.1 motifs ± 25 bp regions, represented using log₂ odds ratio values (two-sided Fisher test, adjusted p values < 0.1). Motifs are summarized as TF families and the specific TF with the maximum absolute log₂ odds ratio value is stated in parenthesis. When TF family annotation was not available, the log₂ odds ratio of the specific TF itself is used. **(C)** Average of DNA shape features at IL-4-induced PU.1 regions of AAM^{mono} and AAM^{res}. Scatter plots are centered on IL-4-induced PU.1 motifs (x -axes), with the solid (Figure legend continues)

in accessibility counts between C57BL/6 and BALB/c at FDR 10%) or strain-common (no significant differences in accessibility count between C57BL/6 and BALB/c) IL-4-induced PU.1 regions from both AAM^{mono} and AAM^{res}. We next focused only on strain-specific IL-4-induced PU.1 regions with SNP(s) flanking the PU.1 motif within ± 25 bp (Supplemental Fig. 3A). This allowed us to ask if the differences in chromatin accessibility between strains were only due to the flanking sequence mutation(s) and not confounded by disruption of a potential PU.1 binding site. We then correlated these strain-specific differences in accessibility with phenotypic differences in gene expression.

We highlight this DNA shape dependent regulatory mechanism at two independent genomic regions that we infer to be regulating biologically relevant functions. Our first example is an intergenic accessible region in AAM^{res}, in which two different SNPs occurring at the 19th and 24th base pair upstream of a PU.1 motif led to changes in DNA shape, most drastically with the ProT and HelT configurations (Fig. 4A). This was accompanied by decrease in chromatin accessibility counts (\log_2 fold change BALB/c versus C57BL/6 = -2.57 ; adjusted p value = 0.01) and reduced expression of the nearby *Slc30a4* gene in the BALB/c mice (\log_2 fold change BALB/c versus C57BL/6 = -3.43 ; adjusted p value = 0.006). *Slc30a4* is a zinc exporter that affects the survival of *Histoplasma capsulatum* in IL-4-stimulated macrophages (48). Our second example is an intronic region on the *P4hb* gene in AAM^{mono}, in which SNPs occurring at the 8th and 24th base pair upstream of a PU.1 motif led to changes in DNA shape, with the former causing larger shifts in the MGW and roll configurations (Fig. 4B). This was, however, accompanied by increased accessibility instead (\log_2 fold change BALB/c versus C57BL/6 = 1.87, adjusted p value = 0.0085) and increased expression of the *P4hb* gene (\log_2 fold change BALB/c versus C57BL/6 = 0.34, adjusted p value = 0.048).

Although the PU.1 motifs were intact in the regions with disrupted accessibility described above, the flanking SNPs could have disrupted the motif of a cofactor to cause the alterations in PU.1 accessibility. Because our experiments are performed on peritoneal macrophages of different cellular lineages isolated ex vivo from mice treated with IL-4 in vivo, it is difficult to obtain sufficient cell numbers from mice directly for multiple TF ChIP-seq experiments. Fortunately, a recent study systematically compared TF binding and function with macrophages from five different strains of mice to assess the impact of genetic variation on TFs activity in macrophages (43). Hence, we could analyze multiple published TF ChIP-seq (including PU.1, CEBP, cJUN, RUNX, and USF) and ATAC-seq datasets generated from nonstimulated BMDMs of C57BL/6 and SPRET/Ei genetic backgrounds. From the non-stimulated BMDMs of C57BL/6 genetic background, we first identified a set of 673 accessible PU.1 motif sites that were bound by PU.1 alone, but not by any of the other cofactors tested in the study. We compared the binding of these 673 PU.1 motif sites between nonstimulated BMDMs of C57BL/6 and SPRET/Ei genetic backgrounds to identify 67 motif sites that contained SNPs and had reduced PU.1 binding. Of these 67 SNP-containing PU.1 binding sites with strain-specific binding pattern, $\sim 40\%$ (26 of 67) had SNPs that flanked the intact PU.1 motif. An example of such a PU.1 binding event occurred at the first exon of the gene *RGMB* (Fig. 4C), which has been described as a binding partner of PD-L2

(49). Although *RGMB* was expressed in BMDMs of C57BL/6, it was not detected in BMDMs of SPRET/Ei background (Fig. 4C). Hence, we can identify SNPs in flanking regions to an intact PU.1 motif, which can disrupt PU.1 binding in macrophages of different genetic backgrounds.

Overall, we found that PU.1 motif-flanking variants occurred at higher frequency than variants located directly within the motif (24.3% versus 5.4% in AAM^{res} and 20.7% versus 6.9% in AAM^{mono}) (Fig. 5A). This suggests that DNA shape readout (a proxy for local environment), instead of direct base readout during PU.1 binding, is likely to be a more important sequence property that could influence PU.1 binding. The frequency of local variants was also greater in strain-specific genomic elements when compared with strain-common regions (two-sided Fisher exact test; $p = 4.59 \times 10^{-4}$ in AAM^{res}; $p = 8.12 \times 10^{-6}$ in AAM^{mono}), even though strain-common PU.1 motifs could also have sequence variants present within ± 25 bp of the PU.1 motif (Fig. 5B). When we extended this analysis to all accessible regions in IL-4 stimulated macrophages, regardless of the presence of PU.1 motif or IL-4 inducibility, we see that 17.5% of all strain-common regions carry local variants. This suggests that the presence of sequence variants do not always have an impact on chromatin accessibility and additional factors must influence whether accessibility at specific chromatin regions are more likely to be altered by sequence variants.

Given the potential role of DNA shape in regulating chromatin accessibility, we first asked if the amount of DNA shape change caused by a sequence variant was associated with its impact on chromatin accessibility. We identified all accessible regions carrying SNP(s), regardless of the presence of a PU.1 motif or IL-4 inducibility and categorized them into either strain-specific or strain-common, as defined above (Supplemental Fig. 3B). The strain-specific regions had a significantly greater change in DNA shape secondary to sequence variants and this was most significant in the MGW, ProT, and roll DNA shape configurations (Fig. 5C). We next used a logistic regression model to systematically examine the contribution of various DNA sequence features in altering chromatin accessibility (detailed in *Materials and Methods*). We found IL-4 inducibility, SNP frequency, and presence of a nearby CpG island to be significant predictors of strain-specific accessible regions (Fig. 5D). Accessible regions that were IL-4 inducible and regions that contained higher frequency of SNPs were more likely to be strain specific, whereas accessible regions overlapping a CpG island were more likely to be strain common. Consistent with this finding, strain-specific regions (regardless of the presence of variants) were significantly more enriched with IL-4-induced peaks overall, and this was observed in both AAM^{mono} and AAM^{res} (Fig. 5E). This suggests that differences in the accessible chromatin landscape of activated macrophages on different genetics background were more likely to be driven by stimulation-dependent genomic elements, further supporting the functionality of IL-4-induced genomic elements during alternative activation. We also asked if the higher frequency of SNPs in a local chromatin region could have contributed to alteration of chromatin accessibility by inducing a greater change in DNA shape. Indeed, SNP frequency positively correlated with the amount of SNP-induced DNA shape change, especially for the

lines representing Loess fit of predicted shape values at single-nucleotide resolution (dots). Boxplots are average predicted shape values over the PU.1 motif ± 25 bp windows, and p values are from two-sided Kolmogorov-Smirnov test. (D) Predicted DNA shape at the eighth base pair of PU.1 motif of AAM^{res} and AAM^{mono}. Frequency distributions are represented by smoothed kernel density estimates. HelT, helical twist; MGW, minor groove width; ProT, propeller twist.

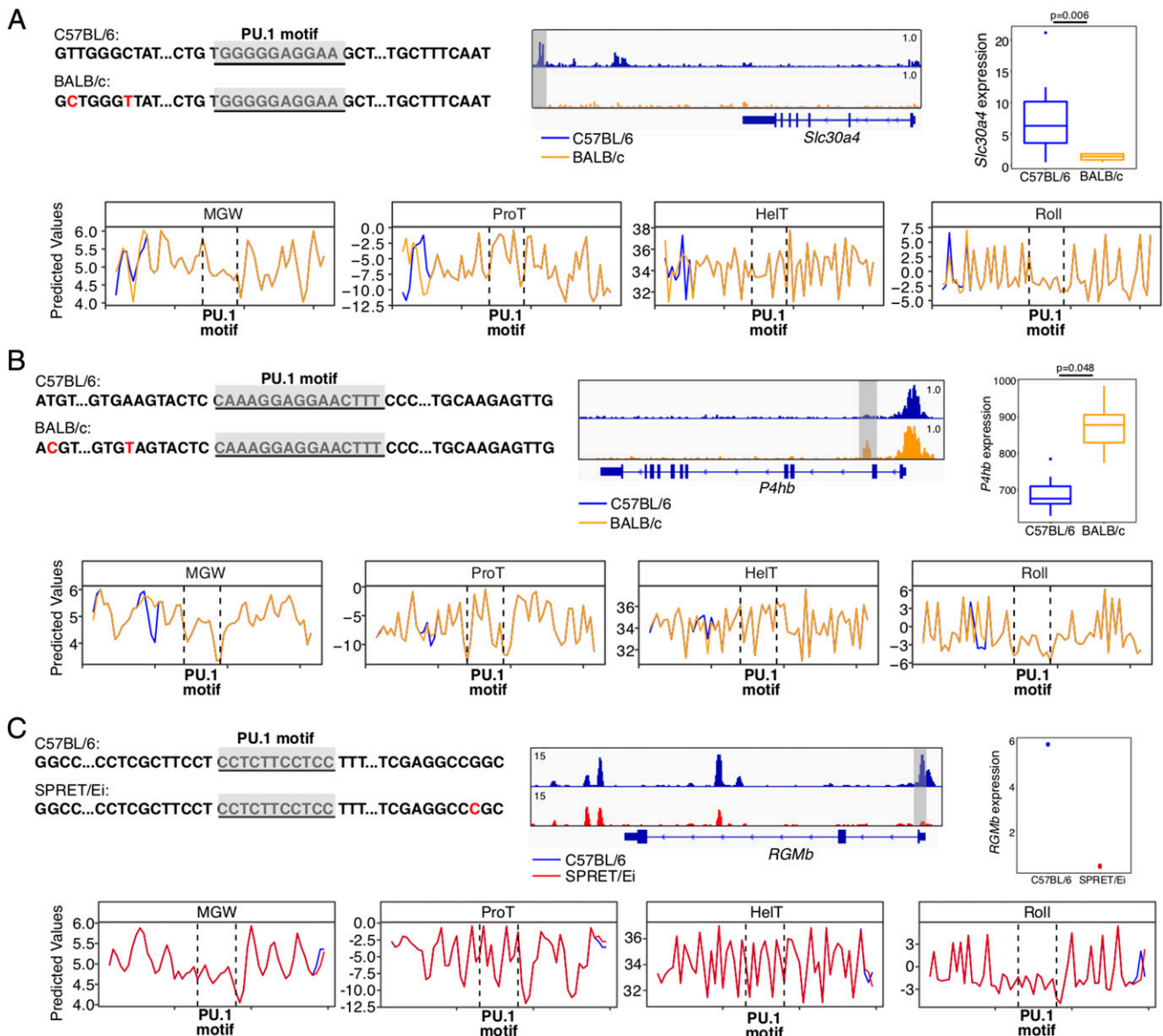


FIGURE 4. Accessibility of IL-4-induced PU.1 regions can be altered through DNA shape change induced by sequence mutation without disruption of the PU.1 motif. Genomic regions containing PU.1 motif with flanking SNPs, selected from (A) AAM^{res}, (B) AAM^{mono}, and (C) nonstimulated BMDMs from Ref. 43. In all examples, sequences represent a unique PU.1 motif region, in which the PU.1 motif is underlined and shaded in gray, whereas SNPs are highlighted with red font. Line graphs of predicted shape values are centered on the PU.1 motif (sequences between the vertical lines). Genome browser tracks illustrate the strain-specific accessibility of the region containing the matching sequence (shaded in gray). ATAC-seq reads are visualized in (A) and (B), whereas ChIP-seq reads are visualized in (C). Boxplots in (A) and (B) represent size factor–normalized read counts, with *p* values are from DESeq2 (*n* = 4–8 mice per group) and adjusted by Benjamini–Hochberg procedure. *RGMb* expression in (C) is expressed as transcript per kilobase million, with values directly obtained from Ref. 43.

shape features MGW and ProT (Fig. 5F). Overall, these results provide support for our hypothesis that SNPs between C57BL/6 and BALB/c that cause alterations in predicted local DNA shape features contribute toward changes in chromatin accessibility in response to IL-4 stimulation.

AAMs from C57BL/6 and BALB/c mice are functionally distinct

We next compared the global transcriptional profiles of AAM^{res} and AAM^{mono} from C57BL/6 and BALB/c mice to identify functional differences in these AAMs of different genetic backgrounds. Even though most of the differences in transcriptional profiles were driven by the cellular origin of macrophages (45% of total variance), strain differences also contributed to the considerable variation in transcriptional profiles (18% of total variance) (Fig. 6A).

Consistent with this finding, most of the macrophage-specific functions were conserved across mouse strains and not affected by genetic differences (Fig. 6B, left panel). In contrast, functional differences secondary to genetics were largely specific to the different types of macrophages (Fig. 6B, right panel).

We next examined the strain-specific functional differences in AAM^{res} and AAM^{mono}, respectively (Fig. 6C). BALB/c AAM^{res} expressed lower levels of cell cycle–related genes, in line with the previously reported observation that peritoneal AAM^{res} have lower proliferation capacity during *Litomosoides sigmodontis* infection in BALB/c mice (10). Furthermore, the expressions of PD-L2 (*Pdcd1lg2*) (Fig. 6D) and MHCII molecules (Fig. 6E), which are cellular markers typically used to characterize alternative activation (12) in AAM^{mono} of C57BL/6 background, were

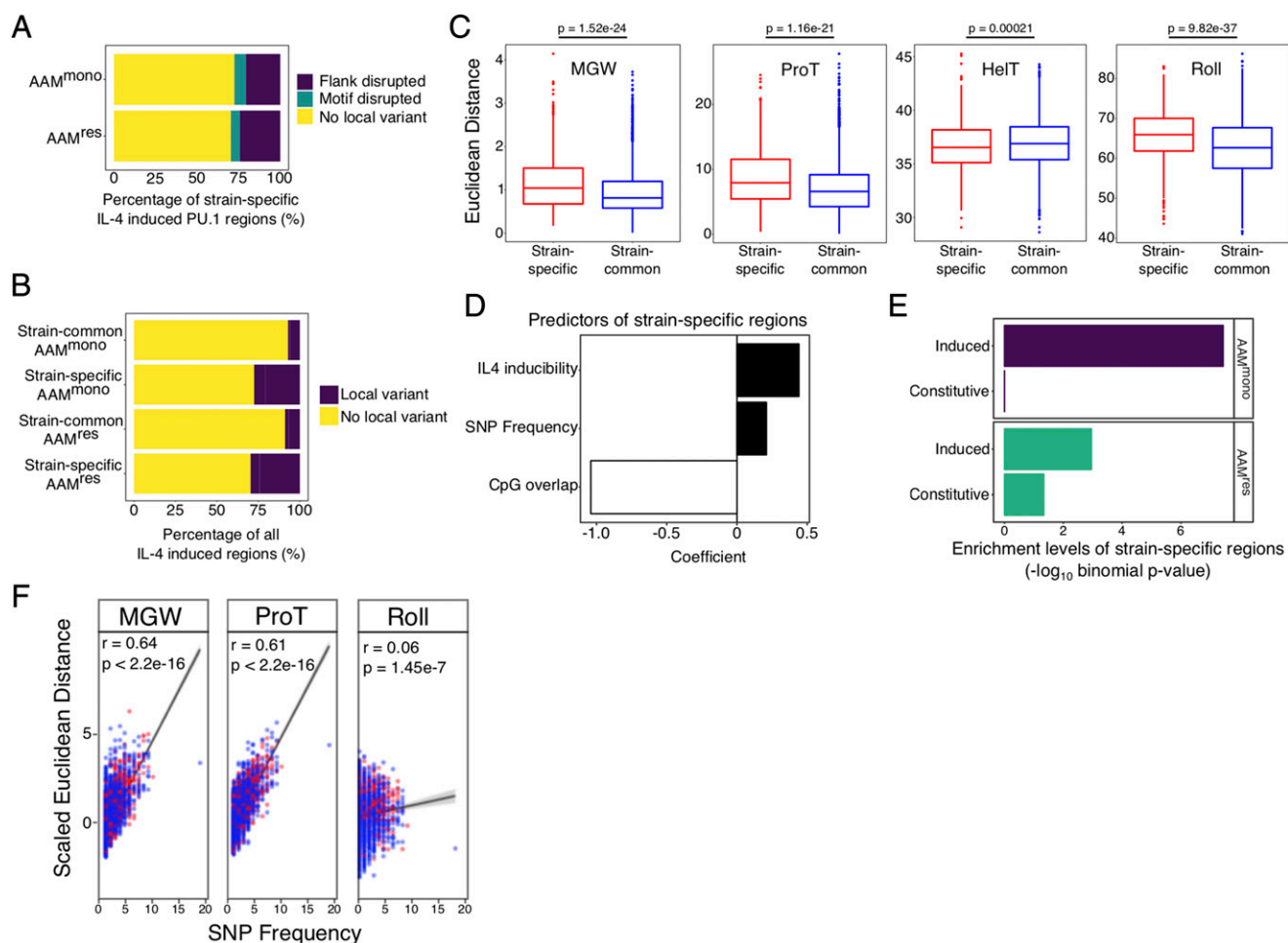


FIGURE 5. Effects of SNP variants on chromatin accessibility can be predicted by the extent of DNA shape change and association with IL-4 inducibility. **(A)** Frequency of different types of sequence variants in strain-specific IL-4-induced PU.1 regions and **(B)** all IL-4-induced PU.1 regions (bottom). Number of strain-specific IL-4-induced PU.1 regions = 58 in AAM^{mono} and 37 in AAM^{res}; number of strain-common IL-4-induced PU.1 regions = 754 in AAM^{mono} and 828 in AAM^{res}. **(C)** Comparison of the amount of SNP-induced DNA shape change in 922 SNP(s)-containing strain-specific and 4776 SNP(s)-containing strain-common accessible regions, respectively. The shift in DNA shape for each sequence is quantitated using Euclidean distance, in which a larger Euclidean distance value indicates a larger shift in shape. The *p* values are from Mann–Whitney tests. **(D)** Coefficient values representing the relative contribution of each significant predictor in distinguishing between strain-specific and strain-common accessible regions. **(E)** Enrichment values of IL-4-induced and constitutively accessible regions in strain-specific regions for AAM^{mono} and AAM^{res}. Enrichment *p* value is based on binomial test. **(F)** Scatter plots illustrating the positive correlations between SNP frequency and amount of DNA shape change undergone by a specific region. Red, strain-specific regions; blue, strain-common regions. Euclidean distances in this study are scaled across each DNA shape feature for visualization purpose. HelT not shown as not correlated. Correlation coefficient and *p* values are based on the Spearman test using unscaled Euclidean distances.

significantly reduced in BALB/c AAM^{mono}. Notably, although three of the five MHCII genes (*H2-Aa*, *H2-Ab1*, and *H2-Eb1*) had significantly higher expression in C57BL/6 AAM^{mono}, *H2-Ea-ps* expression was specific to BALB/c AAM^{mono}. Hence, our studies, together with published findings (10), indicate that macrophages from BALB/c and C57BL/6 mice are functionally distinct in how they respond to IL-4 activation. The SNPs between C57BL/6 and BALB/c mice that contribute toward changes in chromatin accessibility in regulatory regions may also contribute toward these functional differences in macrophage phenotypes. However, further work is needed to demonstrate a link between alterations in predicted local DNA shape features around the PU.1 motifs and the different functionality of macrophages between the C57BL/6 and BALB/c mice.

Discussion

In this study, we define and characterize IL-4-induced chromatin accessibility with in vivo alternative activation of tissue-resident and monocyte-derived peritoneal macrophages. Although the

pioneer factor PU.1 motif is similarly associated with IL-4-induced regions in both macrophage lineages, the local DNA shape features flanking these PU.1 motifs depend on the cellular origin of the macrophages. By leveraging natural genetic variation between mouse strains, we demonstrate that the accessibility of putative PU.1 binding sites correlates with the presence of SNPs flanking the PU.1 binding site, local DNA shape predictions, and altered expression of proximal genes. For alternative activation by IL-4, systematic investigations into the molecular basis of differential activation in different strains of mice are lacking. Although our initial goal was to explore how different lineages of macrophages (tissue resident versus inflammatory monocyte derived) remodel their chromatin in response to the same IL-4 stimuli, our results led us to examine how specific DNA shape features may contribute toward chromatin remodeling around binding motifs for the pioneer factor PU.1. In turn, this led us to test our model using natural genetic variations between BALB/c and C57BL/6 mice. An exciting possibility from our studies is that DNA sequence shape features could be an important contributor toward

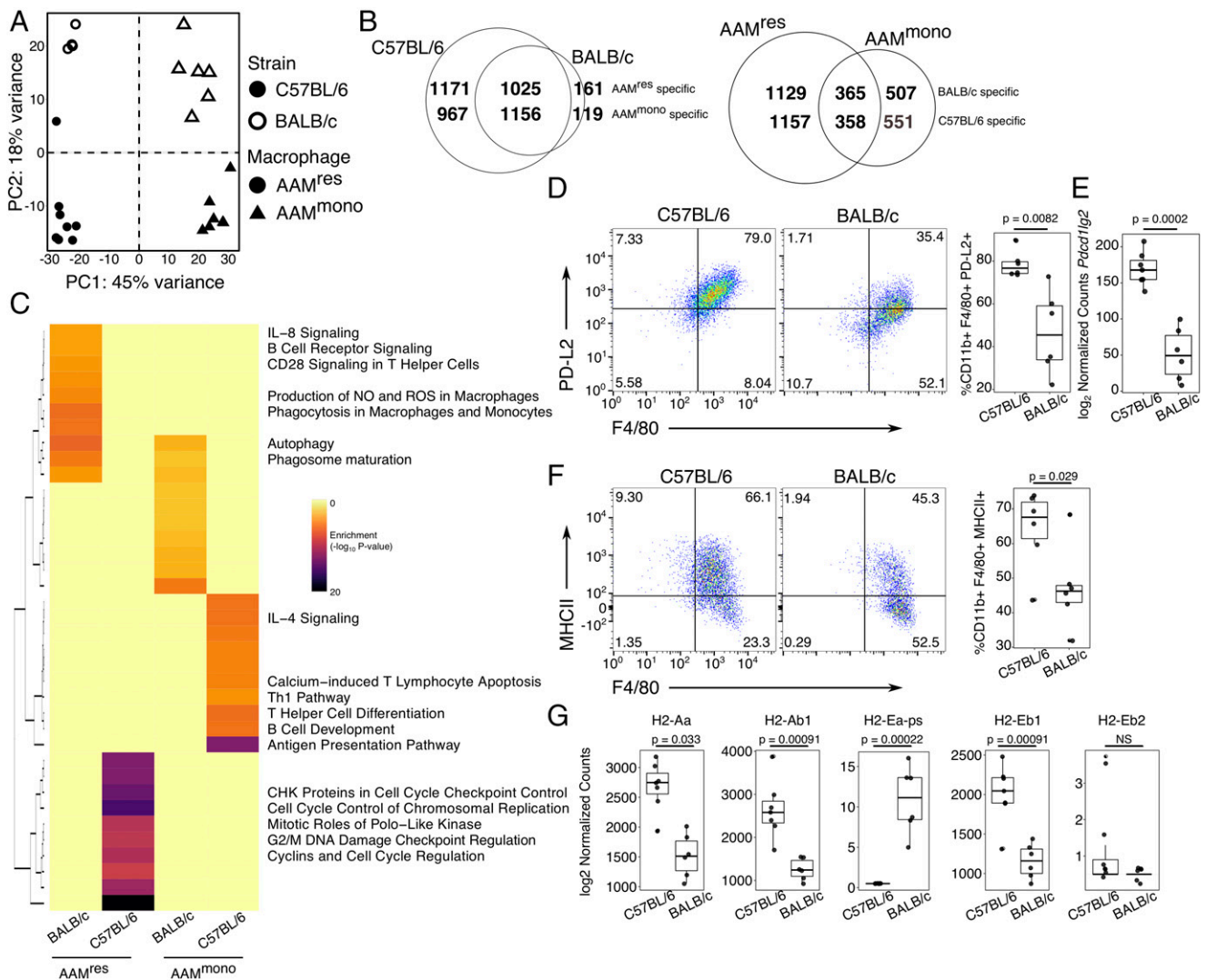


FIGURE 6. AAMs from C57BL/6 and BALB/c are functionally distinct. **(A)** PCA of 7431 genes with high variance (only genes with variance interquartile range >0.5 were retained). Data points represent independent biological replicates. **(B)** Venn diagrams indicating the number of genes that were commonly and uniquely identified as significantly differential (FDR < 0.1) in different comparisons—(left) macrophage-specific genes in C57BL/6 and BALB/c AAMs and (right) strain-specific genes in AAM^{res} and AAM^{mono}. **(C)** Enrichment values from Ingenuity Pathway Analysis visualized as $-\log_{10} p$ value for the four different groups of genes—1) BALB/c specific in AAM^{res}, 2) C57BL/6 specific in AAM^{res}, 3) BALB/c specific in AAM^{mono}, and 4) C57BL/6 in AAM^{mono}. Only the top 10 pathways (as defined by enrichment p values) are included in this visualization. Specific pathways are highlighted for clarity. **(D)** Representative flow cytometric analysis of F4/80 and PD-L2 surface expressions in AAM^{mono} of C57BL/6 and BALB/c mice. Boxplots show frequency of CD11b⁺ F4/80⁺ PD-L2⁺ singlet, live cells. The p value is based on a two-sided unpaired t test. **(E)** Expression of the *Pdcd1lg2* gene in AAM^{mono} of C57BL/6 versus BALB/c mice, represented by size-factor normalized read counts. The p value is from DESeq2 and adjusted by Benjamini-Hochberg procedure. **(F)** Representative flow cytometric analysis of F4/80 and MHCII surface expressions in AAM^{mono} of C57BL/6 and BALB/c mice. Boxplots show frequency of CD11b⁺ F4/80⁺ MHCII⁺ singlet, live cells. The p value is based on a two-sided unpaired t test. **(G)** Expression values of all MHCII genes in AAM^{mono} of C57BL/6 versus BALB/c mice, represented by size-factor normalized read counts. The p values are from DESeq2 and adjusted by Benjamini-Hochberg procedure. Hinges of all boxplots correspond to values of the 25th, 50th, and 75th percentiles, whereas boxplot whiskers extend to no more than $1.5 \times$ interquartile range, beyond which the outlier data points will be plotted individually. Transcriptional profiling analysis: $n = 8$ AAM^{res} (C57BL/6), 4 AAM^{res} (BALB/c), 7 AAM^{mono} (C57BL/6), and 6 AAM^{mono} (BALB/c). Flow cytometric analysis: $n = 6$ AAM^{mono} (C57BL/6) and 6 AAM^{mono} (BALB/c).

explaining the variation in TF binding, gene expression, and chromatin state during macrophage activation in the context of IL-4 activation.

From our findings, we propose that PU.1 is one of the key regulators of IL-4-induced chromatin accessibility and PU.1 binding can be mediated through DNA shape readout of the flanking sequences. DNA shape features, particularly the DNA minor groove width and roll configuration, have been used to distinguish between functional PU.1 binding sites and randomly occurring PU.1 motif (50). However, because DNA shape is a

consequence of DNA sequence, both modes of DNA recognition are confounded and difficult to dissociate from one another. For example, it is unclear if the differences of DNA shape in the PU.1 regions of AAM^{res} and AAM^{mono} are simply due to differences in cofactor binding, or if the PU.1 protein in these different lineages of macrophages have different posttranslational modifications and recognize binding sites with different DNA shape. Therefore, it is important for future studies to identify which amino acid residue(s) in the PU.1 protein could be involved with DNA shape readout, as mutating these residues could potentially be a strategy

to identify PU.1 binding sites that are solely dependent on DNA shape readout, without being confounded by DNA sequences and binding of transcriptional cofactors (51).

Many critical questions remain unanswered. Do these de novo accessible regions persist after resolution of the external stimulation to cause differences in activation kinetics with a repeated stimulation, or would they lose accessibility to allow for plasticity in macrophage activation? What is the specific relationship between the differences in chromatin remodeling between the C57BL/6 and BALB/C macrophages and their distinct functional differences at baseline or after secondary repeated stimulation? In addition, what are the intermediate molecular events that could explain cell type-specific local DNA shape at PU.1 regions? Possible mechanisms include differences in specific TFs cooperatively binding with PU.1, different protein-protein interactions occurring at different functional domains of the PU.1 protein in a macrophage-specific manner, or variants of the PU.1 protein secondary to different posttranslational modifications of PU.1. Our statistical power for examining the contribution of variants to cell type-specific local DNA shape at PU.1 regions would also have been greatly enhanced with the inclusion of PWK or SPRET mice.

The peritoneal IL-4c injection model that we use in this study was chosen as a system to work with primary macrophages isolated from a well-defined tissue environment, whereby the cellular origins of the macrophages have been carefully characterized by multiple laboratories. Although these peritoneal macrophages may not be as physiologically relevant as those generated during a natural parasitic helminth infection (e.g., *Heligomosomoides polygyrus* or *Litomosomoides sigmodontis*), we consider this a stepping-stone between in vitro BMDMs and a physiological immune response during infection or inflammation. Future experiments will determine if we can also use natural genetic variation between different strains of mice to carefully dissect the chromatin remodeling and transcriptional regulation of helminth induced macrophages during infection. Furthermore, our studies in this work were limited by ATAC-seq analysis alone and ChIP-seq data on histone modifications, as well as TFs such as PU.1 and STAT6 would provide additional evidence as to whether there is a strong link between alterations in predicted local DNA shape features, TF binding in enhancer or other regulatory elements and the functionality of macrophages between the C57BL/6 and BALB/c mice.

Because trait-associated variants in the human population are concentrated in noncoding regulatory DNA and disruption of these regulatory elements could be a mechanism of how variants lead to specific phenotypes, it is possible that DNA shape change could be a means to predict which regulatory element are more likely to be disrupted by sequence variants, leading to phenotypic predictions in human health and disease. Hence, future work to understand how DNA shape features could mechanistically control the complex relationship between chromatin structure and gene expression during macrophage activation may provide new insights into how noncoding genetic variation influences cellular phenotypes during inflammatory conditions in human diseases.

Acknowledgments

We thank the NYU School of Medicine Genome Technology Center and Cytometry and Cell Sorting Laboratory for the use of core facilities and the NYU IT High Performance Computing for its resources, services, and staff expertise. We thank Verena Link for providing intermediate files that enabled the analysis reported in Fig. 4C.

Disclosures

The authors have no financial conflicts of interest.

References

- Glass, C. K., and G. Natoli. 2016. Molecular control of activation and priming in macrophages. *Nat. Immunol.* 17: 26–33.
- Ghisletti, S., I. Barozzi, F. Mietton, S. Polletti, F. De Santa, E. Venturini, L. Gregory, L. Lonie, A. Chew, C.-L. Wei, et al. 2010. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* 32: 317–328.
- Natoli, G., S. Ghisletti, and I. Barozzi. 2011. The genomic landscapes of inflammation. *Genes Dev.* 25: 101–106.
- Ostuni, R., V. Piccolo, I. Barozzi, S. Polletti, A. Termanini, S. Bonifacio, A. Curina, E. Prosperini, S. Ghisletti, and G. Natoli. 2013. Latent enhancers activated by stimulation in differentiated cells. *Cell* 152: 157–171.
- Kaikkonen, M. U., N. J. Spann, S. Heinz, C. E. Romanoski, K. A. Allison, J. D. Stender, H. B. Chun, D. F. Tough, R. K. Prinjha, C. Benner, and C. K. Glass. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* 51: 310–325.
- Comoglio, F., M. Simonatto, S. Polletti, X. Liu, S. T. Smale, I. Barozzi, and G. Natoli. 2019. Dissection of acute stimulus-inducible nucleosome remodeling in mammalian cells. *Genes Dev.* 33: 1159–1174.
- Herbert, D. R., C. Hölscher, M. Mohrs, B. Arendse, A. Schwegmann, M. Radwanska, M. Leeto, R. Kirsch, P. Hall, H. Mossmann, et al. 2004. Alternative macrophage activation is essential for survival during schistosomiasis and downmodulates T helper 1 responses and immunopathology. [Published erratum appears in 2004 *Immunity* 21: 455.] *Immunity* 20: 623–635.
- Chen, F., W. Wu, A. Millman, J. F. Craft, E. Chen, N. Patel, J. L. Boucher, J. F. Urban, Jr., C. C. Kim, and W. C. Gause. 2014. Neutrophils prime a long-lived effector macrophage phenotype that mediates accelerated helminth expulsion. *Nat. Immunol.* 15: 938–946.
- Chen, F., Z. Liu, W. Wu, C. Roza, S. Bowdridge, A. Millman, N. Van Rooijen, J. F. Urban, Jr., T. A. Wynn, and W. C. Gause. 2012. An essential role for TH2-type responses in limiting acute tissue damage during experimental helminth infection. *Nat. Med.* 18: 260–266.
- Campbell, S. M., J. A. Knipper, D. Ruckerl, C. M. Finlay, N. Logan, C. M. Minutti, M. Mack, S. J. Jenkins, M. D. Taylor, and J. E. Allen. 2018. Myeloid cell recruitment versus local proliferation differentiates susceptibility from resistance to filarial infection. *eLife* 7: e30947.
- Gundra, U. M., N. M. Girgis, M. A. Gonzalez, M. San Tang, H. J. P. Van Der Zande, J.-D. Lin, M. Ouimet, L. J. Ma, J. Poles, N. Vozhilla, et al. 2017. Vitamin A mediates conversion of monocyte-derived macrophages into tissue-resident macrophages during alternative activation. *Nat. Immunol.* 18: 642–653.
- Gundra, U. M., N. M. Girgis, D. Ruckerl, S. Jenkins, L. N. Ward, Z. D. Kurtz, K. E. Wiens, M. S. Tang, U. Basu-Roy, A. Mansukhani, et al. 2014. Alternatively activated macrophages derived from monocytes and tissue macrophages are phenotypically and functionally distinct. *Blood* 123: e110–e122.
- Daniel, B., G. Nagy, A. Horvath, Z. Czimmerer, I. Cuaranta-Monroy, S. Poliska, T. T. Hays, S. Sauer, J. Francois-Deleuze, and L. Nagy. 2018. The IL-4/STAT6/PPAR γ signaling axis is driving the expansion of the RXR heterodimer cistrome, providing complex ligand responsiveness in macrophages. *Nucleic Acids Res.* 46: 4425–4439.
- Daniel, B., G. Nagy, Z. Czimmerer, A. Horvath, D. W. Hammers, I. Cuaranta-Monroy, S. Poliska, P. Tzerpos, Z. Kolostyak, T. T. Hays, et al. 2018. The nuclear receptor PPAR γ controls progressive macrophage polarization as a ligand-insensitive epigenomic ratchet of transcriptional memory. *Immunity* 49: 615–626.e6.
- Czimmerer, Z., B. Daniel, A. Horvath, D. Ruckerl, G. Nagy, M. Kiss, M. Pelloquin, M. M. Budai, I. Cuaranta-Monroy, Z. Simandi, et al. 2018. The transcription factor STAT6 mediates direct repression of inflammatory enhancers and limits activation of alternatively polarized macrophages. *Immunity* 48: 75–90.e6.
- Jenkins, S. J., D. Ruckerl, P. C. Cook, L. H. Jones, F. D. Finkelman, N. van Rooijen, A. S. MacDonald, and J. E. Allen. 2011. Local macrophage proliferation, rather than recruitment from the blood, is a signature of TH2 inflammation. *Science* 332: 1284–1288.
- Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10: 1213–1218.
- Hashimshony, T., N. Senderovich, G. Avital, A. Klochendler, Y. de Leeuw, L. Anavy, D. Gennert, S. Li, K. J. Livak, O. Rozenblatt-Rosen, et al. 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* 17: 77.
- Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Breese, M. R., and Y. Liu. 2013. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics* 29: 494–496.
- Thorvaldsdóttir, H., J. T. Robinson, and J. P. Mesirov. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14: 178–192.
- McCarthy, M. T., and C. A. O'Callaghan. 2014. PeakDECK: a kernel density estimator-based peak calling program for DNase-seq data. *Bioinformatics* 30: 1302–1304.
- Stark, R., and G. Brown. 2011. DiffBind: differential binding analysis of ChIP-Seq peak data. Available at: <http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf>. Accessed: July 1, 2017.
- Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15: 550.

26. Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43: e47.
27. Gentleman, R., V. Carey, W. Huber, and F. Hahne. 2016. genefilter: methods for filtering genes from high-throughput experiments. R package version 1.56.0 ed.
28. Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
29. Hsu, F., W. J. Kent, H. Clawson, R. M. Kuhn, M. Diekhans, and D. Haussler. 2006. The UCSC known genes. *Bioinformatics* 22: 1036–1046.
30. Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, and W. S. Noble. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37(Web Server issue): W202–8.
31. Karwacz, K., E. R. Miraldi, M. Pokrovskii, A. Madi, N. Yosef, I. Wortman, X. Chen, A. Watters, N. Carriero, A. Awasthi, et al. 2017. Critical role of IRF1 and BATF in forming chromatin landscape during type 1 regulatory cell differentiation. *Nat. Immunol.* 18: 412–421.
32. Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38: 576–589.
33. Mahony, S., P. E. Auron, and P. V. Benos. 2007. DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies. *PLOS Comput. Biol.* 3: e61.
34. Mahony, S., and P. V. Benos. 2007. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res.* 35(Web Server issue): W253–8.
35. Heinz, S., C. E. Romanoski, C. Benner, K. A. Allison, M. U. Kaikkonen, L. D. Orozco, and C. K. Glass. 2013. Effect of natural genetic variation on enhancer selection and function. *Nature* 503: 487–492.
36. Grant, C. E., T. L. Bailey, and W. S. Noble. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27: 1017–1018.
37. Wingender, E., T. Schoeps, M. Haubrock, M. Krull, and J. Dönitz. 2018. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* 46(D1): D343–D347.
38. Chiu, T.-P., F. Comoglio, T. Zhou, L. Yang, R. Paro, and R. Rohs. 2016. DNashapeR: an R/Bioconductor package for DNA shape prediction and feature encoding. *Bioinformatics* 32: 1211–1213.
39. Zhou, T., L. Yang, Y. Lu, I. Dror, A. C. Dantas Machado, T. Ghane, R. Di Felice, and R. Rohs. 2013. DNashape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.* 41(Web Server issue): W56–62.
40. Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.
41. Neph, S., M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman, A. K. Johnson, E. Rynes, M. T. Maurano, J. Vierstra, S. Thomas, et al. 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28: 1919–1920.
42. Rozowsky, J., A. Abyzov, J. Wang, P. Alves, D. Raha, A. Harmanci, J. Leng, R. Bjornson, Y. Kong, N. Kitabayashi, et al. 2011. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* 7: 522.
43. Link, V. M., S. H. Duttker, H. B. Chun, I. R. Holtman, E. Westin, M. A. Hoeksema, Y. Abe, D. Skola, C. E. Romanoski, J. Tao, et al. 2018. Analysis of genetically diverse macrophages reveals local and domain-wide mechanisms that control transcription factor binding and function. *Cell* 173: 1796–1809. e1717.
44. Rosas, M., L. C. Davies, P. J. Giles, C. T. Liao, B. Kharfan, T. C. Stone, V. B. O'Donnell, D. J. Fraser, S. A. Jones, and P. R. Taylor. 2014. The transcription factor Gata6 links tissue macrophage phenotype and proliferative renewal. *Science* 344: 645–648.
45. Jarjour, N. N., E. A. Schwarzkopf, T. R. Bradstreet, I. Shchukina, C. C. Lin, S. C. Huang, C. W. Lai, M. E. Cook, R. Taneja, T. S. Stappenbeck, et al. 2019. Bhlhe40 mediates tissue-specific control of macrophage proliferation in homeostasis and type 2 immunity. *Nat. Immunol.* 20: 687–700.
46. Dror, I., T. Golan, C. Levy, R. Rohs, and Y. Mandel-Gutfreund. 2015. A widespread role of the motif environment in transcription factor binding across diverse protein families. *Genome Res.* 25: 1268–1280.
47. Gordán, R., N. Shen, I. Dror, T. Zhou, J. Horton, R. Rohs, and M. L. Bulyk. 2013. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* 3: 1093–1104.
48. George, M. M., K. Subramanian Vignesh, J. A. Landero Figueroa, J. A. Caruso, and G. S. Deepe. 2016. Zinc induces dendritic cell tolerogenic phenotype and skews regulatory T cell–Th17 balance. *J. Immunol.* 197: 1864–1876.
49. Xiao, Y., S. Yu, B. Zhu, D. Bedoret, X. Bu, L. M. Francisco, P. Hua, J. S. Duke-Cohan, D. T. Umetsu, A. H. Sharpe, et al. 2014. RGMb is a novel binding partner for PD-L2 and its engagement with PD-L2 promotes respiratory tolerance. *J. Exp. Med.* 211: 943–959.
50. Barozzi, I., M. Simonatto, S. Bonifacio, L. Yang, R. Rohs, S. Ghisletti, and G. Natoli. 2014. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol. Cell* 54: 844–857.
51. Abe, N., I. Dror, L. Yang, M. Slatery, T. Zhou, H. J. Bussemaker, R. Rohs, and R. S. Mann. 2015. Deconvolving the recognition of DNA shape from sequence. *Cell* 161: 307–318.